STAT 611 Part 1
Basics

Zepu Zhang
May 4, 2011

1 Expectation, variance, covariance, correlation

(Appendix A of Chapter 2)

Assume continuous random variables. Discrete ones need some adjustment to the formulas; basically, integrals become sums.

X, Y, Z are variables. a, b, c are constants.

1.1 Definition and properties of expectation

Univariate definition:

$$E\big(h(X)\big) = \int h(x)\, f(x)\, \mathrm{d}x$$

In particular, if $h(x) \equiv x$, then $E(X) = \int x\, f(x)\, \mathrm{d}x$.

Multivariate definition:

$$E\big(h(X,Y)\big) = \iint h(x,y)\, f(x,y)\, \mathrm{d}x\, \mathrm{d}y$$

The most important property of expectation is its linearity:

$$E(aX + b) = aE(X) + b$$
$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

1.2 Definition and properties of variance

Definition:

$$\mathrm{var}(X) = E\Big[\big(X - E(x)\big)^2\Big]$$

Properties:

$$\mathrm{var}(aX + b) = a^2\, \mathrm{var}(X)$$

(Scaling gets squared; shifting does not change variance.)

$$\mathrm{var}(X) = E(X^2) - \big(E(X)\big)^2$$

(Very useful for computations.)

1.3 Definition and properties of covariance

Definition:

$$\mathrm{cov}(X, Y) = E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big]$$

Properties:

$$\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$$
$$\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y)$$
$$\mathrm{cov}(X, c) = 0$$
$$\mathrm{cov}(X, cY) = c\,\mathrm{cov}(X, Y)$$
$$\mathrm{cov}(X + Y, Z) = \mathrm{cov}(X, Z) + \mathrm{cov}(Y, Z)$$

Note: X, Y, Z here do not have to be a simple random variable. For example, Z may be $W + 3U$. Using the above we get

$$\mathrm{cov}(aX + b, cY + d) = ab\,\mathrm{cov}(X, Y)$$

$$\mathrm{cov}(X_1 + \cdots + X_m, Y_1 + \cdots + Y_n) = \sum_{i=1}^{m} \sum_{j=1}^{n} \mathrm{cov}(X_i, Y_j)$$

Note $\mathrm{var}(X) \equiv \mathrm{cov}(X, X)$ (verified directly by the definitions). Consequently, some properties of variance can be obtained from properties of covariances:

$$\mathrm{var}(a_1X_1 + \cdots + a_nX_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_ia_j\,\mathrm{cov}(X_i, X_j)$$

$$\mathrm{var}(X+Y) = \mathrm{cov}(X, X)+\mathrm{cov}(X, Y)+\mathrm{cov}(Y, X)+\mathrm{cov}(Y, Y) = \mathrm{var}(X)+\mathrm{var}(Y)+2\,\mathrm{cov}(X, Y)$$

If X and Y are independent, then

$$\mathrm{cov}(X, Y) = E\big[\big(X-E(X)\big)\big(Y-E(Y)\big)\big] = E\big(X-E(X)\big)\,E\big(Y-E(Y)\big) = 0,$$

hence $\mathrm{var}(X+Y) = \mathrm{var}(X)+\mathrm{var}(Y)$. This generalizes straightforwardly to the sum of multiple independent random variables.

1.4 Definition and properties of correlation

Definition:

$$\mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}}$$

Properties:

$$0 \le |\mathrm{corr}(X, Y)| \le 1$$

i.e.

$$|\mathrm{cov}(X, Y)| \le \sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}$$

1.5 Standardization

$$X^* = \frac{X - E(X)}{\sqrt{\operatorname{var} X}}$$

Properties:

$$E(X^*) = 0, \quad \operatorname{var}(X^*) = 1$$

2 Basic concepts of time series

A stochastic process is a family of random variables $\{X_t\}$, where t is an index. This process is studied by investigating the joint distribution of any finite number of member variables in this family.

To make things tractable, we often do not study the complete joint distributions; instead we only study the first two moments, that is, means, variances, and covariances (or correlations).

Note: if the joint distributions are normal, then the first two moments completely determine the distributions.

If the index is time, in particular, not arbitrary time instants but regular, equi-spaced time units, the stochastic process is a time series concerned in this course.

To follow the text's notation, we'll use Y to denote the random variable. The index t take integer values: 0, ± 1, ± 2,...

We will be investigating the first two moments, that is, $E(Y_t)$, $\operatorname{var}(Y_t)$, $\operatorname{cov}(Y_t, Y_s)$, $\operatorname{corr}(Y_t, Y_s)$. There are functions of the time indices, t and/or s. Often, we will emphasize the time lag and write time indices as t, $t+1$, $t+2$,..., $t+k$, etc.

mean function

autocovariance function (use symbol γ following the text)

autocorrelation function (use symbol ρ following the text)

2.1 Sample autocorrelation function (ACF)

(p. 46)

sample autocorrelation function

correlogram

2.2 stationarity

strictly stationary: joint distribution depends on relative time configuration but not on actual time points.

With strict stationarity, we have (1) mean is constant over time; (2) variance is constant over time; (3) covariance (or correlation) depends on absolute value of time lag but not on actual time location.

weakly stationary: mean is constant over time; covariance depends on absolute value of time lag but not on actual time location.

Since we study the first two moments only, weak stationarity suffices for us. (We won't tell the difference between strict and weak stationarities, because we don't study the parts where they possibly differ.)

If the joint distributions are all normal, then weak and strict stationarities are the same.

2.3 Typical examples

1. White noise
2. Random walk
3. Moving average
4. Cosine wave

Which of these processes are stationary?

Random walk is an accumulation of white noise.

Moving average is moving average of white noise.

Differencing random walk gets white noise.

Cosine wave shows the difficulty in assessing whether stationarity is a reasonable assumption on the basis of the time sequence plot of the observed data.

"Differencing" is an important (and simple) technique from getting stationary series from nonstationary ones.

3 Estimation of a constant mean

(Section 3.2)

Exercise 2.17. (For stationary $\{Y_t\}$.)

Suppose

$$Y_t = \mu + X_t$$

where $E(X_t) = 0$.

Given observations $Y_1, \ldots, Y_n$, take

$$\overline{Y} = \frac{1}{n} \sum_{t=1}^{n} Y_t$$

then

$$E\big(\overline{Y}\big) = \frac{1}{n} \sum_{t=1}^{n} E(Y_t) = \frac{1}{n} \sum_{t=1}^{n} \big(\mu + E(X_t)\big) = \mu$$

therefore $\overline{Y}$ is an unbiased estimator of the constant mean, μ.

Besides the mean (which is good, unbiased), we mainly want to also know the variance, because that tells us how "precise" the estimator tends to be.

To study more properties of $\overline{Y}$, such as its variance, we need to assume more things about Y_t, or equivalently, X_t. Now assume stationarity, then the result of Exercise 2.17 can be used.

$$\mathrm{var}\big(\overline{Y}\big) = \ldots$$

With this formula (for stationary Y_t), we look at a few special cases:

- Moving average: $Y_t = e_t - .5e_{t-1}$.

  How much difference can the sign of ρ_1 make?

- Autocorrelation decaying fast enough such that $\sum_{k=0}^{\infty} |\rho_k| < \infty$, and n is large.

  A general approximation: (3.2.5)

  Example: $\rho_k = \phi^{|k|}$.

For a nonstationary process, the estimator $\overline{Y}$ can be very imprecise (although still unbiased). Example: random walk.